

STOR 151 SECTION 2 MIDTERM 1 FEBRUARY 18 2010

This is an open book exam. Course text, personal notes and calculator are permitted. You have 75 minutes to complete the test. Personal computers and cellphones are not to be used during the exam. If you have any queries about the meaning of the questions, or if you think there is an error, ask the proctor for assistance. Answers are to be written in a blue book.

Before you begin:

- (a) Write your name and PID on the front of your blue book.
- (b) Write your five-digit number [**OPTIONAL**: This should be personal to you. Do not use your PID and do not disclose to any other person, but do keep a record of it. This is for display of results on the webpage. If you don't want your results displayed, leave blank.]
- (c) Sign the "pledge". If this is not preprinted, copy out the following statement and sign it: *On my honor, I have neither given nor received unauthorized aid in this exam.*

SHOW ALL WORKING — even correct answers will not get full credit if it's not clear how they were obtained. Incorrect answers will gain substantial credit if the method of working is substantially correct.

Answer all questions. The score for each part is indicated at the end of the question (total 100).

1. Please read the attached article, "As Girls Become Women, Sports Pay Dividends", by Tara Parker-Pope of the New York Times. The article refers to two studies, one conducted by Dr. Stevenson at the Wharton School and the other by Dr. Kaestner of the University of Illinois at Chicago. This part of the exam will ask you a number of questions about those studies.

The Stevenson study:

- (i) Was it an observational study or an experiment? [4]
- (ii) Was it prospective or retrospective? [4]
- (iii) List *one* response variable, the main explanatory variable, and one lurking variable. [10]
- (iv) Which if the following statements best summarizes how this study avoids the problem of lurking variables? (Choose one of A,B,C,D: no further explanation required.) [7]
 - A. Because the selection of participants was random, we should expect that any lurking variables are evenly distributed among the groups being compared.
 - B. The study was stratified by state, so this allowed the author to correct for lurking variables that vary from one state to another.
 - C. The main factor determining the girls' participation in sports was the previously existing boys' participation rate, and this could not have been affected by lurking variables.
 - D. Because the change in girls' participation rates occurred so quickly, it allowed a direct comparison before and after Title IX which did not depend on lurking variables.
- (v) Nevertheless, it's not necessarily the case that *all possible* lurking variables were accounted for. Give one lurking variable that might still be a problem. [5]

[CONTINUED ON OTHER SIDE. PLEASE TURN PAGE.]

The Kaestner study:

- (vi) This is more like a traditional clinical trial in which there are treatments and controls. What are the treatment group and the control group in this case? [6]
- (vii) In the light of your answer to (vi), what would you say is the principal weakness of this study? [7]

Overall:

- (viii) Write in *one sentence* what you feel is the most important conclusion of this article. [7]

2. [This question is also motivated by the Parker-Pope article but it describes a completely different type of study and is independent of question 1, except for part (iv).]

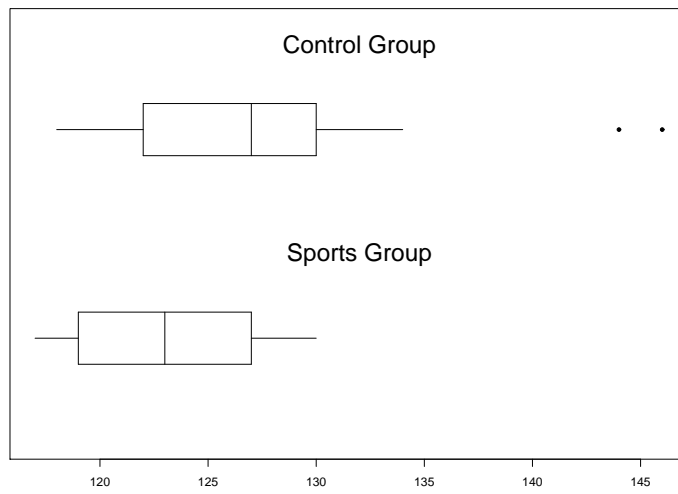
It is proposed to carry out a randomized experiment to study the effect of taking a sports training class on girls' health. A class of 30 female high-school students is divided randomly into two groups of 15. One group is assigned to attend a sports training class, the other (the control group) is assigned to a class that discusses general leisure activities without any emphasis on sports (e.g. music, movies). Both groups are then followed for two years at the end of which a variety of health measures are recorded. In this question we will focus on one of these, the students' weights.

- (i) Assume the students are given two-digit numbers from 01 to 30. Use the table of random numbers given below to show which group of 15 students will be assigned randomly to the sports training class. [*Be specific. Describe exactly how you use the table and give the actual numbers of the students assigned to take the sports class.*] [15]
- (ii) After two years, the weights of the two groups of students are as follows:
Sports group: 126, 128, 130, 123, 117, 118, 123, 120, 127, 121, 129, 118, 126, 119, 124
Control group: 129, 118, 146, 134, 120, 127, 125, 126, 130, 122, 122, 126, 129, 144, 127
Draw a side-by-side pair of boxplots to illustrate the results. [20]
- (iii) Based on your answer to (ii), briefly comment on the differences between the two groups of students. [7]
- (iv) State *one advantage* and *one disadvantage* of a study conducted this way, compared with the kinds of studies mentioned in the Parker-Pope article. [8]

86208508534293066442174120944294171336290338316894351480019723575070068842254384
95015300773516509214583324722433874800118180511379633982693675505108127120329774
30510272429648830154957958859181344149117885599800590755825530300012678861167401
74755401567579295672928241910503055796921386660206275097469300440327392539838237
01118751224869157255520032342463252457168507979686493672969841511283462680899723
56908687633511464497129606869419323174563394930215323031821519703422370375940821
16026805047233897476533706492966626460677742021742267546377348493421354437948991
67754398406216064679553487762092720401042877900166931195831922894993718157232660
91200954913272064133921474588183453541719632902532150834399151137404950804288907
94638299900415273305736706079137997392185076370473482108995792740042955674815201
44877141212426114428382085908461469097128466414018465789462296701698091674404374
31708090119756280306559647964699006850010344144238940756698805692802735040589016
93346242959167753904556677240360259245110201606393455023894750781835278082603256

SKETCH SOLUTIONS

1.
 - (i) Observational.
 - (ii) Retrospective.
 - (iii) Response: Women's education or employment. Explanatory: Participation in sports. Lurking: school size, climate, social and personal differences among athletes.
 - (iv) C. A lot of people answered B and I gave 4/7 for that, but if you read the article carefully, it does explain the critical role played by the variation on boys' participation rate by state.
 - (v) "Participation [still] varies widely by state" — so there are possible lurking variables based on cultural, demographic or economic differences among states. For example, if wealthier states had better pre-existing boys' sports programs, then they would also have been obliged to spend more to bring girls' participation rates up to match the boys' level. But if such states also had better educational and employment outcomes, it would be difficult to state definitively that such differences were due to sports.
 - (vi) The treatment group was women who had been in high school in the 1970s, while the control group was similar women from earlier years.
 - (vii) The study used historical controls. There could be many other reasons for a trend (e.g. improving health care overall) which may have nothing to do with sports. *Alternatively:* It's not a true controlled experiment because there was no randomization of participants to treatment and control.
 - (viii) *Possible answer.* Title IX increased girls' participation rates in sports, and this had benefits for their long-term health, education and employment.
2.
 - (i) A possible answer: begin in row 3 of the random number table, select pairs of numbers deleting any outside the range 01-30 and ignoring duplicates. Continue until 15 are selected. The result: students 30, 02, 01, 11, 07, 12, 16, 29, 05, 03, 13, 06, 27, 25, 22 are selected for the sports class. [Many students chose to start with the first row, in which case the selected students are numbers 20, 08, 06, 17, 13, 29, 03, 14, 01, 23, 25, 16, 24, 11, 12.]
 - (ii) Sports group: five-number summary is 117, 119, 123, 127, 130; outlier limits are 107, 139. Control group: five-number summary is 118, 122, 127, 130, 146; outlier limits are 110, 142 (so 144, 146 are outliers).
Boxplots: See next page.
 - (iii) The sports group has a slightly lower overall weight and also a more symmetric distribution. The control group is right-skewed and has two outliers. Nevertheless, there is still substantial overlap between the two distributions.
 - (iv) The studies mentioned in the Parker-Pope article were not randomized studies and therefore may still be affected by lurking variables. However, this study also has a significant weakness — it's based on a small sample size and consequently there is not a clear-cut conclusion that the two groups are really different.



Boxplots for question 2(ii).